

Hierarchical Multivariate Space-Time Methods for Modeling Counts with an Application to Stroke Mortality Data

Harrison Quick^{*1}, Lance A. Waller², and Michele Casper¹

¹ Division of Heart Disease and Stroke Prevention, Centers for Disease Control and Prevention, Atlanta, GA 30329

² Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322

* *email:* HQuick@cdc.gov

SUMMARY. Geographic patterns in stroke mortality have been studied as far back as the 1960s, when a region of the southeastern United States became known as the “stroke belt” due to its unusually high rates of stroke mortality. While stroke mortality rates are known to increase exponentially with age, an investigation of spatiotemporal trends by age group at the county-level is daunting due to the preponderance of small population sizes and/or few stroke events by age group. Here, we harness the power of a complex, nonseparable multivariate space-time model which borrows strength across space, time, and age group to obtain reliable estimates of yearly county-level mortality rates from US counties between 1973 and 2013 for those aged 65+. Furthermore, we propose an alternative metric for measuring changes in event rates over time which accounts for the full trajectory of a county’s event rates, as opposed to simply comparing the rates at the beginning and end of the study period. In our analysis of the stroke data, we identify differing spatiotemporal trends in mortality rates across age groups, shed light on the gains achieved in the Deep South, and provide evidence that a separable model is inappropriate for these data.

KEY WORDS: Age disparities in health, Bayesian methods, Nonseparable models, Saved person-years, Small area analysis

1 Introduction

Stroke (i.e., cerebrovascular disease) is the fourth leading cause of death in the general population and third — behind heart disease and cancer — among those aged 85 and older, with rates increasing exponentially with age (Kochanek et al., 2015). Geographic trends in stroke mortality have been studied as far back as the 1960s, when Borhani (1965) identified a region of the southeastern United States (US) stretching from Mississippi to North Carolina which had the highest rates of stroke mortality — a region which would become known as the “stroke belt.” Later work by Casper et al. (1995) noticed an apparent shift in the stroke belt, observing that regions of the Mississippi River Valley appeared in the highest decile of mortality rates in the early 1990s where they had previously not. More recently, Schieb et al. (2013) have studied geographic trends in stroke hospitalizations from 1995 to 2006, noting that, among Medicare beneficiaries ages 65 and older, this shift in the stroke belt has persisted, stretching further into parts of Texas and Oklahoma. In addition to changing geographic patterns, numerous studies have observed the overall declines in stroke mortality (e.g., Howard et al., 2001; Gillum et al., 2011).

While many of these studies have age-adjusted their data, accounting for the variation in age distributions among counties and disparities in stroke mortality across age ranges, this precludes inference within individual age groups. On the other hand, this aggregation and data standardization step also helps mitigate the issue of small population sizes and low number of stroke deaths in many US counties, an issue that can lead to unreliable rate estimates and is only exacerbated when the data are stratified by a factor such as age group. In this work, however, our goal is to investigate spatiotemporal trends in stroke mortality by jointly modeling data from three age-based subpopulations, permitting more reliable inference at the county level for each age group while preserving the ability to compute age-adjusted rates. More specifically, we look to build a complex multivariate space-time model which borrows strength across space, time, and age group. We also propose a new tool for measuring declines in mortality which accounts for temporal changes in mortality rates.

To achieve this, a natural starting point is the family of models based on the conditionally autoregressive (CAR) model proposed by Besag (1974). Since its extension to the fully Bayesian setting in Besag et al. (1991), CAR models have sparked a wealth of research in the disease mapping context for both spatial (e.g., Besag et al., 1995; Besag and Higdon, 1999) and spatiotemporal applications. While these early examples were all based on the standard univariate CAR model, Gelfand and Vounatsou (2003) developed methods for general multivariate CAR (MCAR) models, inspiring novel approaches for both multiple and spatiotemporal disease mapping (e.g., Jin et al., 2007; Quick et al., 2013, 2015a; Martinez-Beneito, 2013; Botella-Rocamora et al., 2015). More recently, Quick et al. (2015c) proposed a special case of the MCAR of Gelfand and Vounatsou (2003) — referred to as the multivariate space-time CAR (MSTCAR) model — for the purpose of analyzing county-level heart disease mortality rates in the US over time for various race/gender groups. For a more complete coverage of the recent advances in spatial and space-time modeling, see Banerjee et al. (2014).

In any discussion of the spatiotemporal modeling literature, we would be remiss not to mention the subject of *separability* — i.e., models where the spatiotemporal covariance can be decomposed into the product of a purely spatial covariance and a purely temporal covariance. While separable covariance structures offer computational benefits, Stein (2005) highlights some drawbacks of the separability assumption, and concerns over the utility of separable models have motivated the development of classes of nonseparable covariance functions in the univariate continuous space, continuous time setting (e.g., Cressie and Huang, 1999; Gneiting, 2002). In cases where both space and time are discrete, as encountered in this study, Knorr-Held (2000) has discussed a variety of possible space-time interactions, but here again the focus has been on a single outcome, with extensions to the multivariate space-time setting being more recent developments. In addition to the nonseparable MSTCAR proposed by Quick et al. (2015c), Bradley et al. (2014) implement a reduced rank multivariate spatiotemporal mixed effects model which is designed to analyze high dimensional data efficiently. That said, both of these approaches restrict their attention to the case of Gaussian

data. Other methods (e.g., Jin et al., 2007; Martinez-Beneito, 2013) allow for varying spatial structures by utilizing *proper* MCAR models — while these approaches are feasible when the number of spatial regions in the spatial domain is small, the large number of US counties prohibits the use of proper MCAR models.

Here we extend the nonseparable MSTCAR model to the generalized linear model setting to analyze spatiotemporal trends in the dataset comprised of stroke mortality counts among those aged 65–74, 75–84, and 85+ described in Section 2. Specifically, due to the rarity of stroke deaths, the Gaussian assumption of Quick et al. (2015c) may not be appropriate. As such, in Section 3 we detail our approach for embedding the MSTCAR model into a Poisson likelihood akin to Besag et al. (1991), in addition to presenting the saved person-years (SPY) tool for measuring patterns in declines. We then analyze the stroke mortality data in Section 4, where we discover different spatiotemporal trends in mortality rates across age groups and observe evidence that a separable model would be inappropriate for these data. Finally, we summarize our findings and offer some concluding remarks in Section 5.

2 Data Description

The study population for this analysis includes all US residents aged 65 or older. In order to assess differences across the high-risk age ranges, we separate our data into $N_g = 3$ groups: those aged 65–74, those 75–84, and those 85+. Annual counts of stroke-related deaths per county per age group were obtained from the National Vital Statistics System (NVSS) of the National Center for Health Statistics (NCHS). Due to inconsistencies in the manner in which death records were recorded prior to 1973, we restrict the analysis to data from 1973–2013 ($N_t = 41$ years) to ensure valid comparisons across time. Deaths from stroke were defined as those for which the underlying cause of death was cerebrovascular disease according to the 8th, 9th, and 10th revisions of the International Classification of Diseases (ICD; ICD-8: 430–438; ICD-9: 430–438; ICD-10: I60–69). Based on the comparability ratios reported by Klebba and

Scott (1980) and Anderson et al. (2001), which indicate a high degree of similarity between the three revisions of the ICD, we assume that this definition is consistent over the 41 year study period. The geographic unit used in this analysis was the county (or county equivalent). Given changes in county definitions during the study period affecting ten counties (e.g., the merging/splitting of counties), a single set of $N_s = 3,099$ regions (henceforth referred to as counties) from the contiguous lower 48 states (including the District of Columbia) was used for the entire study period. Annual population counts were based on the bridged-race intercensal estimates provided by NCHS (2013). While the number of individuals in each age bracket has risen considerably (by 89%, 101%, and 267%, respectively), the number of stroke-related deaths for individuals 65–84 has decreased nearly 60% since 1973. From a public health perspective, this reduction in stroke mortality is a great achievement. From a statistical perspective, however, this can lead to concerns over the reliability of county-level mortality rate estimates based on so few events. Figures illustrating the national population and death trends for each age group can be found in Web Appendix B.

3 Methods

3.1 Statistical model

Letting Y_i and n_i denote the incidence of disease and the population at risk in county i , Besag et al. (1991) proposed a model of the form

$$Y_i \sim \text{Pois}(n_i \exp[\mathbf{x}_i \boldsymbol{\beta} + Z_i + \phi_i]), \text{ for } i = 1, \dots, N_s \quad (1)$$

where \mathbf{x}_i denotes a p -vector of covariates with corresponding regression coefficients, $\boldsymbol{\beta}$, Z_i is a spatial random effect, and $\phi_i \stackrel{\text{ind}}{\sim} N(0, \tau^2)$ is an exchangeable random effect. In their work, Besag et al. (1991) modeled $\mathbf{Z} = (Z_1, \dots, Z_{N_s})'$ as arising from an intrinsic conditional

autoregressive (CAR) model, which has the conditional distribution

$$Z_i | \mathbf{Z}_{(i)}, \sigma^2 \sim N \left(\sum_{j=1}^{N_s} w_{ij} Z_j / \sum_{j=1}^{N_s} w_{ij}, \sigma^2 / \sum_{j=1}^{N_s} w_{ij} \right) \quad (2)$$

where $\mathbf{Z}_{(i)}$ denotes the vector \mathbf{Z} with the i th element removed and $w_{ij} = 1$ if i and j are neighbors (denoted $i \sim j$) and 0 otherwise. Recommendations for prior distributions for σ^2 and τ^2 are offered by Bernardinelli et al. (1995) and Waller et al. (1997).

Extending (1) and (2) to a setting consisting of multiple spatial surfaces is straightforward. Letting Y_{ikt} denote the number of deaths in county i during year t for age group k , we model

$$Y_{ikt} \sim \text{Pois} (n_{ikt} \exp [\mathbf{x}_{ikt} \boldsymbol{\beta}_{kt} + Z_{ikt} + \phi_{ikt}]) , \quad (3)$$

for $i = 1, \dots, N_s$, $k = 1, \dots, N_g$, and $t = 1, \dots, N_t$ where $\phi_{ikt} \stackrel{\text{ind}}{\sim} N(0, \tau_k^2)$. To account for the multivariate spatiotemporal association in the data, we follow the MSTCAR model of Quick et al. (2015c) — itself a special case of the MCAR of Gelfand and Vounatsou (2003) — and let $\mathbf{Z} = (\mathbf{Z}'_{1..}, \dots, \mathbf{Z}'_{N_s..})' \sim \text{MCAR}(1, \Sigma_\eta)$,

$$\pi(\mathbf{Z} | \Sigma_\eta) \propto |\Sigma_\eta|^{-(N_s-1)/2} \exp \left[-\frac{1}{2} \mathbf{Z}' \{ (D - W) \otimes \Sigma_\eta^{-1} \} \mathbf{Z} \right]$$

and $\mathbf{Z}_{i..} | \mathbf{Z}_{(i)..}, \Sigma_\eta \sim N \left(\sum_{j \sim i} \mathbf{Z}_{j..} / m_i, \frac{1}{m_i} \Sigma_\eta \right),$

where $\mathbf{Z}_{i..} = (\mathbf{Z}_{i1..}, \dots, \mathbf{Z}_{iN_g..})'$, $\mathbf{Z}_{ik.} = (Z_{ik1}, \dots, Z_{ikN_t})'$, W is an adjacency matrix with elements w_{ij} , D is an $N_s \times N_s$ diagonal matrix with elements $m_i = \sum_{j=1}^{N_s} w_{ij}$, Σ_η denotes the $N_t N_g \times N_t N_g$ covariance structure for our N_t years and N_g age groups and \otimes denotes the Kronecker product. The η subscript is a reference to the construction of \mathbf{Z} from Quick et al. (2015c) where the authors began by defining $\mathbf{v}_{\iota t} \stackrel{\text{iid}}{\sim} N(\mathbf{0}, G_t)$, where $\iota = 1, \dots, N_s - 1$, to be a collection of independent N_g -dimensional random variables with covariance G_t for $\iota = 1, \dots, (N_s - 1)$ and $t = 1, \dots, N_t$. Using $R_k = R(\rho_k)$ to denote an age group-specific

temporal correlation matrix based on an autoregressive order 1 (denoted AR(1)) model and letting \tilde{R}_k to be the Cholesky decomposition of R_k such that $\tilde{R}_k \tilde{R}_k' = R_k$, the authors define $\boldsymbol{\eta}_{\iota k \cdot} = \tilde{R}_k \mathbf{v}_{\iota k \cdot}$ where $\mathbf{v}_{\iota k \cdot} = (v_{\iota k 1}, \dots, v_{\iota k N_t})'$. We then find that $\boldsymbol{\eta}_{\iota \cdot} \sim N(\mathbf{0}, \Sigma_\eta)$, where

$$\Sigma_\eta = \begin{bmatrix} \tilde{R}_{1,1}^* & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \mathbf{0} \\ \tilde{R}_{N_t,1}^* & \cdots & \tilde{R}_{N_t,N_t}^* \end{bmatrix} \begin{bmatrix} G_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & G_{N_t} \end{bmatrix} \begin{bmatrix} \tilde{R}_{1,1}^* & \cdots & \tilde{R}_{N_t,1}^* \\ \mathbf{0} & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \tilde{R}_{N_t,N_t}^* \end{bmatrix} \quad (4)$$

and $\tilde{R}_{t,t'}^*$ denotes the $N_g \times N_g$ diagonal matrix with elements $\{\tilde{R}_k\}_{t,t'}$ for $k = 1, \dots, N_g$; \mathbf{Z} is then constructed from the $\boldsymbol{\eta}_{\iota \cdot}$ using the eigenvalues and eigenvectors of the matrix $D - W$ (see Rue and Held, 2005). This structure is then denoted $\mathbf{Z} \sim \text{MSTCAR}(G_1, \dots, G_{N_t}, \boldsymbol{\rho})$.

3.2 Hierarchical model and computational details

While a Poisson model like (3) is a straightforward extension of (1), such models can also pose computational challenges, particularly for large dimensions. For instance, the full conditional of $\mathbf{Z}_{i \cdot}$, given by

$$\begin{aligned} \pi(\mathbf{Z}_{i \cdot} \mid \mathbf{Y}, \mathbf{Z}_{(i) \cdot}, \boldsymbol{\beta}, \boldsymbol{\phi}, \Sigma_\eta) &\propto \prod_{k=1}^{N_g} \prod_{t=1}^{N_t} \text{Pois}(Y_{ikt} \mid n_{ikt} \exp[\mathbf{x}_{ikt} \boldsymbol{\beta}_{kt} + Z_{ikt} + \phi_{ikt}]) \\ &\times \pi(\mathbf{Z}_{i \cdot} \mid \mathbf{Z}_{(i) \cdot}, \Sigma_\eta) \end{aligned} \quad (5)$$

is *not* a known distribution. That is, if we use a Markov chain Monte Carlo (MCMC) algorithm to estimate the posterior distribution of our model parameters, this model may require the use of large multivariate Metropolis updates within our Gibbs sampler. Besag et al. (1995) and Knorr-Held and Rue (2002) suggest a reparameterization of (3) which involves integrating ϕ_{ikt} out of the model, yielding a Gaussian full conditional for $\mathbf{Z}_{i \cdot}$ and requiring Metropolis updates for $\theta_{ikt} = \mathbf{x}_{ikt} \boldsymbol{\beta}_{kt} + Z_{ikt} + \phi_{ikt}$. Fortunately, the θ_{ikt} are independent of one another given \mathbf{Y} and the other model parameters, so these Metropolis updates can be

conducted independently and in parallel.

We complete our hierarchical model by specifying the following prior distributions for our other model parameters: a vague prior for β , a weakly informative inverse Gamma prior for each τ_k^2 , a beta prior for each ρ_k , and an inverse Wishart prior for each G_t with hyperparameter G , itself modeled using a Wishart prior. While this structure on the covariance matrices is likely unnecessary given the number of spatial regions in the data (see the discussion of prior sensitivity in spatial models by Bernardinelli et al., 1995), this comes at little-to-no computational cost (see Web Appendix A.6) and offers a convenient means for specifying proper priors. Putting these pieces together, our full hierarchical model is as follows:

$$\begin{aligned} \pi(\beta, \mathbf{Z}, G, G_1, \dots, G_t, \rho, \tau_k^2, \theta | \mathbf{Y}) &\propto \prod_{i,k,t} \text{Pois}(Y_{ikt} | n_{ikt} \exp[\theta_{ikt}]) \times N(\theta | X\beta + \mathbf{Z}, \Sigma_\theta) \\ &\times \text{MSTCAR}(\mathbf{Z} | G_1, \dots, G_{N_t}, \rho) \\ &\times \prod_{t=1}^{N_t} \text{InvWish}(G_t | G, \nu) \times \text{Wish}(G | G_0, \nu_0) \\ &\times \prod_{k=1}^{N_g} [\text{Beta}(\rho_k | a_\rho, b_\rho) \times \text{IG}(\tau_k^2 | a_\tau, b_\tau)], \end{aligned} \quad (6)$$

where Σ_θ is a diagonal matrix of size $N_s N_g N_t$ with elements τ_k^2 and X is the $(N_s N_g N_t \times p)$ matrix of covariates.

While full details for implementing this model in an MCMC framework are provided in Web Appendix A, we would be remiss to not discuss the computational burden associated with fitting a nonseparable model as opposed to a separable model; i.e., letting $\rho_k = \rho$ for $k = 1, \dots, N_g$ and $G_t = G$ for $t = 1, \dots, N_t$ corresponds to fitting a separable multivariate space-time model with $\Sigma_\eta = R(\rho) \otimes G$. First note that by using an AR(1) model for time, we can compute the $\tilde{R}_{t,t'}^*$ elements of Σ_η in closed-form, reducing the burden of computing Σ_η^{-1} from an $N_t N_g \times N_t N_g$ matrix inversion to a series of $N_g \times N_g$ matrix inversions. Furthermore, while the nonseparable MSTCAR model contains more parameters than its separable counterpart, the additional computational burden associated with its

implementation in an MCMC framework is negligible. Specifically, the computations necessary to construct the full conditional distributions for each G_t are simply a partitioning of those necessary for constructing the full conditional distribution of G in a separable model.

When implementing this model in an MCMC framework, we have found that proper specification of the initial values can be crucial to facilitating convergence in a timely fashion. In particular, the parameters which require Metropolis updates — ρ_k and θ_{ikt} — should be treated with care. For instance, we recommend initializing ρ_k to be large (say, 0.90) if a high degree of temporal correlation is expected. More importantly, we recommend initializing

$$\theta_{ikt}^{(0)} = \begin{cases} \log(Y_{ikt}/n_{ikt}), & \text{if } Y_{ikt} > \epsilon \\ \log(\sum_i Y_{ikt}/\sum_i n_{ikt}), & \text{if } Y_{ikt} \leq \epsilon \end{cases}, \quad (7)$$

where $\epsilon \geq 0$ is some small nonnegative integer, as this will allow the model to learn about parameters such as β and \mathbf{Z} early on in the process; in practice, letting $\epsilon = 0$ has been sufficient. When θ_{ikt} is poorly initialized, however, the MCMC algorithm may take a large number of iterations to recover, resulting in a chain which is slow to converge.

3.3 Assessment of reliability

The primary motivation for this work is to achieve more reliable age-specific mortality rate estimates for these data. In order to assess the reliability of our estimates, we will begin by generating *synthetic* death counts from the posterior predictive distribution for Y_{ikt}^* ,

$$Y_{ikt}^{*(\ell)} | \theta_{ikt}^{(\ell)} \sim \text{Pois} \left(n_{ikt} \exp \left[\theta_{ikt}^{(\ell)} \right] \right), \text{ for } i = 1, \dots, N_s, j = 1, \dots, N_t, k = 1, \dots, N_g,$$

and for $\ell = 1, \dots, L$, where $\theta_{ikt}^{(\ell)}$ denotes the ℓ th (post-burn-in) sample from the posterior distribution of θ_{ikt} . We can then compute the 95% CI for Y_{ikt}^* from these posterior predictions and determine the proportion of each county's $N_t N_g$ 95% CIs that contain the observed Y_{ikt} to estimate the coverage probability. As a baseline for comparison purposes, we will compare

our results to those generated from an empirical Bayesian Poisson-gamma model of the form

$$\boldsymbol{\gamma} \mid \mathbf{Y}, a_{kt}, b_{kt} \propto \prod_{ikt} \text{Pois}(Y_{ikt} \mid n_{ikt}\gamma_{ikt}) \times \text{Gamma}(\gamma_{ikt} \mid a_{kt}, b_{kt}) \quad (8)$$

where $a_{kt} = \sum_i Y_{ikt} / \sum_i n_{ikt} \times b_{kt}$ and $b_{kt} = 1000$, indicating a prior distribution equivalent to 1,000 additional persons with an event rate equal to the national average. We can then generate synthetic death counts from the resulting posterior predictive distribution, $Y_{ikt}^{\dagger(\ell)} \mid \gamma_{ikt}^{(\ell)} \sim \text{Pois}(n_{ikt}\gamma_{ikt}^{(\ell)})$ as before. The method which yields coverage probabilities near 0.95 more consistently will be deemed more reliable.

3.4 Tools for measuring temporal changes in mortality

When studying temporarily-varying mortality rates, it is often of interest to measure the decline from the beginning of the study period to the end. Letting $\lambda_{ikt} = \exp[\theta_{ikt}]$ denote the county-specific mortality rate for group k at time t and letting $\Delta_{ik}(t, t') = (\lambda_{ikt} - \lambda_{ikt'}) / \lambda_{ikt}$ denote the percent change from time t to $t' > t$ for group k in county i , an obvious choice would be to compute $\Delta_{ik}(1, N_t)$ for each county and each group. Similarly, one could define $\Delta_i(t, t') = (\sum_k n_{ikt}\lambda_{ikt} - \sum_k n_{ikt'}\lambda_{ikt'}) / \sum_k n_{ikt}\lambda_{ikt}$ and compute $\Delta_i(1, N_t)$ as an estimate of the county's percent decline.

The drawback of these quantities is that they only account for the rates at the beginning and end of the study period, ignoring the intervening periods. Here, we propose a measure we refer to as the “saved person-years” — or SPY — which can be computed as

$$\text{SPY}_i = \frac{1}{N_t - 1} \sum_{t=2}^{N_t} \frac{\sum_k n_{ikt} (\lambda_{ik1} [1 - \Delta_{.k}(1, t)] - \lambda_{ikt})}{\sum_k n_{ikt}} \times 100,000, \quad (9)$$

where $\Delta_{.k}(1, t) = (\lambda_{.k1} - \lambda_{.kN_t}) / \lambda_{.k1}$ denotes the nationwide average decline from time 1 to time t with $\lambda_{.kt} = \sum_i n_{ikt}\lambda_{ikt} / \sum_i n_{ikt}$. Note that SPY_i is essentially a measure of the deviation between the expected rate for county i at time t if the county had declined at the

rate of the national average — i.e., $\lambda_{ik1} [1 - \Delta_k(1, t)]$ — and the rate we estimate from the model. While this quantity should not replace the crude measure of the percent decline, it is a simple and easily interpretable investigative tool that tells a more thorough story of a region’s trajectory over the study period, as we will demonstrate in our analysis.

In practice, we estimate θ_{ikt} by obtaining samples from its posterior distribution — say, $\theta_{ikt}^{(\ell)}$ for iteration ℓ of our MCMC algorithm. As such, both $\Delta_i(t, t')$ and SPY_i can be computed for each iteration of our MCMC algorithm, resulting in posterior distributions for each of these quantities. From this point forward, however, we restrict our attention to the posterior medians of $\Delta_i(t, t')$ and SPY_i unless otherwise stated.

4 Analysis of the Stroke Mortality Data

In the absence of covariate information, we fit the hierarchical model in (6) to the stroke mortality data described in Section 2 using an intercept term for each combination of year and age group (as required when using an improper CAR model, per Besag and Kooperberg, 1995), forcing the random effects to account for a substantial amount of the spatio-temporal variability in the data. We place an informative beta prior on each ρ_k to encourage higher temporal correlations in the model, and we use a vague inverse Wishart prior for each of the G_t . When running the MCMC algorithm, we thinned our posterior samples for θ_{ikt} by removing 9 out of 10 samples — while this is not theoretically necessary, it reduced the burden of storing excess samples for our nearly 400,000 random effects. Estimates provided are based on posterior medians, and 95% credible intervals (95% CI) were obtained by taking the 2.5- and 97.5-percentiles from the thinned post-burn-in samples. Additional figures, including animations displaying temporal evolutions in the geographic trends, can be found in Web Appendix B.

Before delving into the epidemiologic findings, we evaluate some of the numerous variance parameters permitted by the use of a nonseparable model. While the MSTCAR model in

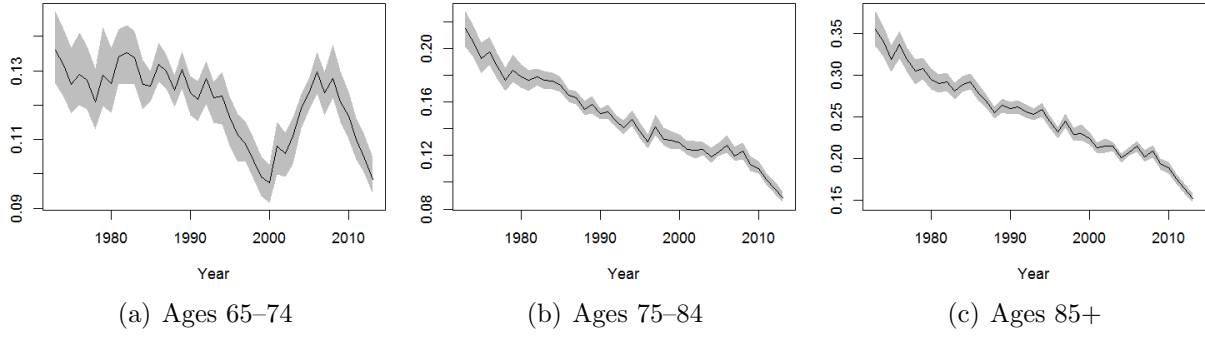


Figure 1: Temporal evolution of the diagonal elements of Σ_η . While the scale of these parameters is not directly interpretable, declines in these variance parameters suggest an increase in spatial smoothing over time (particularly for those 75+).

Section 3.1 is derived using temporally-varying covariance matrices, G_t , these parameters are not necessarily of direct interest as they are the variance parameters for $\mathbf{v}_{\ell,t}$, and thus they are *not* directly interpretable with respect to the mortality rates. Instead, we need to use the posterior samples of G_t and ρ_k to construct Σ_η from (4). These values coincide with the conditional covariance matrix of $\mathbf{Z}_{i\cdot}$ (when scaled by the number of neighbors, m_i), and thus *are* interpretable with respect to the log mortality rates. Furthermore, patterns in these parameters can be easily interpreted, as well. For instance, Figure 1 displays the posterior estimates for the diagonal elements of Σ_η corresponding to each age group. Here, the declines in Figures 1(b) and 1(c) suggest a higher degree of spatial smoothing in later years than at the beginning of the study period, as the corresponding Z_{ikt} become less free to deviate from their neighbors over time. As we will see, this may be in part due to declines in the mortality rates, themselves, as lower rates may also imply a smaller range of rates.

To assess the reliability of our estimates, we follow the approach set forth in Section 3.3. We begin by generating $L = 1000$ replicates for each Y_{ikt} from the posterior predictive distribution corresponding to both the MSTCAR model fit in (6), as well as the empirical Bayesian Poisson-gamma model from (8). After computing the 95% CI of the replicates from both models and comparing these intervals to the observed Y_{ikt} , we find that the mean county-specific coverage from the MSTCAR is 97.7%, whereas the Poisson-gamma yields a

coverage of 99.6%. What is immediately clear from this result is that while the intervals from the MSTCAR are consistently more narrow than those from the Poisson-gamma model, this increase in precision due to the MSTCAR model is valid. Figures related to this reliability assessment are included in Web Appendix B.

Having demonstrated the necessity and utility of the MSTCAR model for these data, we now shift our attention to the rate estimates themselves. For the sake of illustration of the temporal trends and the SPY tool, we highlight two counties: one from the heart of the stroke belt (Jefferson County, AL), and one from the opposite side of the country (King County, WA). Figure 2 displays time trends of the estimated rates for these two counties for each of the age groups, along with the national averages. Here, we observe that while these counties exhibit similar trends for the 65–74 age group — where King County consistently outperforms both Jefferson County and the nation as a whole — the trends for the remaining age groups are quite different. This is particularly true for the eldest age group in our study, as Jefferson County experienced such a sharp decline from 1973 to 1990 so as to not only pass King County, but also the national average. This period of consistent declines was followed by stagnant rates through the early 1990s and an *increase* in rates until the early 2000s, leading to the county ending the study period among the worst in the nation (albeit with much lower rates than in the 1970s).

One aspect highlighted by these figures, however, is that computing a simple percent decline does not tell the whole story. For instance, while Jefferson County experienced some degree of increasing rates in each of the three age groups during the late 1990s, simply measuring the age-standardized decline from 1973 to 2013 would overlook the strides the county took during the first half of the study period, when Jefferson County declined at a rate much faster than the national average. The same cannot be said, however, for King County, which underperformed during the first 30 years of the study, despite experiencing an overall decline that outperformed the national average. This discrepancy can be observed by investigating the SPY_i tool for each county. Here, we find that Jefferson County saved

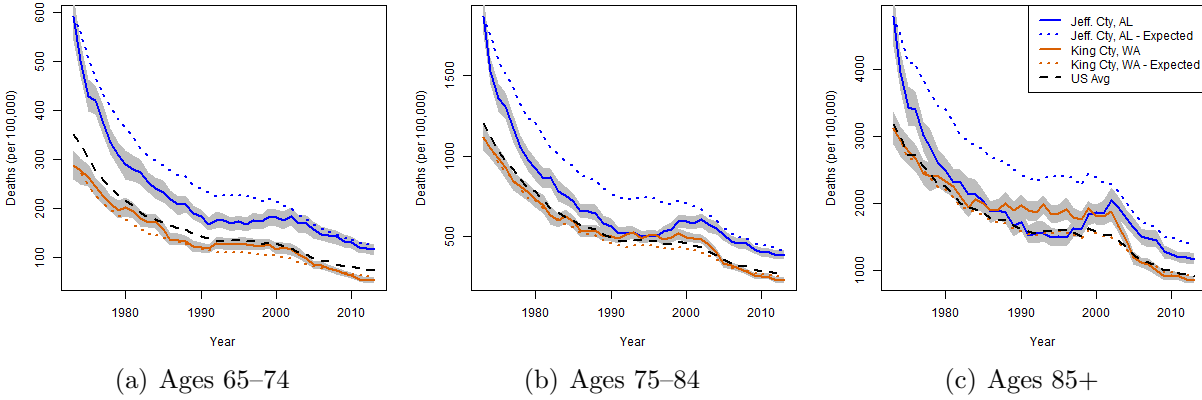


Figure 2: Comparison of the rate trends over time for two selected counties (solid lines), along with the national average (dashed line). Gray bands denote 95% CI. Also displayed for each county is the expected rate had the county declined at the same rate as the national average (dotted lines).

136.6 (98.8, 174.9) person-years per 100,000, while the SPY_i for King County was -34.0 (-61.7, -5.3) person-years per 100,000, reinforcing our claim that King County’s declines lagged behind the national average. A map of the SPY_i values for all 3,099 counties can be found in Figure 3, where we find that parts of the Deep South outperformed much of the nation, and a comparison to the percent decline can be found in Figure B.4 of the Web Appendix.

Turning our attention to the geographic patterns in stroke death rates presented in Figure 4, we find substantial differences between age groups. For the youngest subpopulation (ages 65–74), the clear geographic pattern shown in Figure 4(a) prominently highlights the so-called “stroke belt” in the rates from 1973, and the map of the percent declines in Figure 4(d) — with large declines along the East Coast and smaller declines in the region stretching from Texas to the Dakotas — seems to indicate the “shift” in the stroke belt identified by Casper et al. (1995). These patterns, however, are much less evident among the older age groups, especially for the eldest population. Here, the rates in Figure 4(c) exhibit far less spatial clustering, while Figure 4(f) suggests that the rate of declines in mortality for those 85 and older was generally slower nationwide compared to those observed for those aged 65–74 and 75–84.

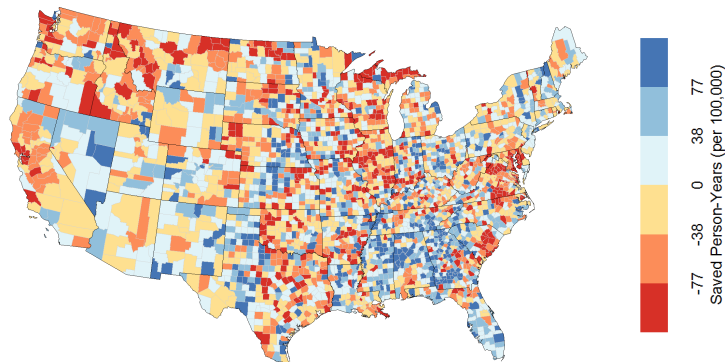


Figure 3: Map of the saved person-years (SPY) measure from (9), which measures the average difference between the model-estimated mortality rate of a county (per 100,000) and the expected mortality rate assuming a rate of decline equivalent to the national average.

5 Discussion

This paper has extended the MSTCAR model of Quick et al. (2015c) to a generalized linear model for the purposes of analyzing a dataset comprised of county-level counts of stroke mortality, allowing for a nonseparable multivariate spatiotemporal dependence structure. Our analysis revealed spatiotemporal trends in stroke mortality that varied by age group, in addition to the nationwide reduction in rates previously noted in the literature (e.g., Gillum et al., 2011; Schieb et al., 2013). We also observed differing aspects of the western shift in parts of the South for each age group, as identified in the total population by Casper et al. (1995), and explored the impact of non-linear trends in stroke mortality via the SPY tool.

When modeling event rates for a rare event such as stroke mortality, it is important to leverage as much information as possible to achieve reliable estimates. In addition to incorporating spatial structure into the model — allowing for regional patterns such as the stroke belt to lend support to less populated counties — the MSTCAR accounts for temporal correlation between observations in consecutive years and multivariate dependencies, such as those between observations in different age-brackets. These additional sources of information can be invaluable when dealing with outlying counts, a problem which particularly plagues counties with small population sizes where an increase/decrease of a single death can

dramatically change the observed mortality rate. Overcoming this challenge will be paramount in our future work, where we wish to use the MSTCAR model to analyze racial and gender disparities in heart disease and stroke mortality across various age groups. While there is computational burden associated with implementing the MSTCAR for such large multivariate datasets, these analyses will provide incredible insight into these various disparities.

Another area for future research is understanding the factors that contribute to differential geographic patterns by age group in both the baseline 1973 stroke mortality rates as well as the patterns of declining stroke mortality rates. While it is well known that the risk for stroke increases with age, the spatiotemporal patterns of stroke mortality by age group have not been documented previously. Hypotheses for understanding the observed differential spatio-temporal patterns in stroke mortality by age group include, but are not limited to, the following categories: 1) spatio-temporal differences in the relative contributions of decreasing case fatality rates and incidence rates by age group (e.g., Ergin et al., 2004; El-Saed et al., 2006); 2) differential influence of living conditions (e.g. socioeconomic resources, access to quality health care, access to healthy food and recreational environments, etc.) or changes in those living conditions, on stroke mortality by age group (e.g., Tassone et al., 2009; Lisabeth et al., 2006); or 3) differences in the accuracy of death certificate reporting by age group due to more co-morbidities and competing conditions of death in the older ages.

Lastly, a topic worth discussing is that publicly available data for rare events such as deaths from stroke can be difficult to find due to data privacy issues. While data such as the number of stroke-related deaths are publicly available at the county-level from NCHS (via CDC Wonder), subsets of data with less than 10 events in a geographic region are suppressed beginning in 1989 (CDC, 2003). This results in over 80% of the data points for those aged 65–74 and nearly 70% of the over 380,000 data points used in this analysis being suppressed to the public. To overcome this privacy issue while still preserving utility, Quick et al. (2015b) have explored the risks associated with the generation of synthetic data for rare event data for a single population for a single year of data. As an area of active research, we hope to use

the methodology proposed here to generate reliable synthetic public-use data for small areas which respects the spatial-, temporal-, and multivariate structures in the true data, thereby providing greater access to complete, high quality data.

References

- Anderson, R. N., Miniño, A. M., Hoyert, D. L., and Rosenberg, H. M. (2001). “Comparability of cause of death between ICD–9 and ICD–10: Preliminary Estimates.” *National Vital Statistics Reports*, 49.
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis of Spatial Data*. Boca Raton, FL: Chapman & Hall/CRC.
- Bernardinelli, L., Clayton, D., and Montomoli, C. (1995). “Bayesian estimates of disease maps: How important are priors?” *Statistics in Medicine*, 14, 2411–2431.
- Besag, J. (1974). “Spatial interaction and the statistical analysis of lattice systems (with Discussion).” *Journal of the Royal Statistical Society, Series B*, 36, 192–236.
- Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). “Bayesian computation and stochastic systems (with discussion).” *Statistical Science*, 10, 3–66.
- Besag, J. and Higdon, D. (1999). “Bayesian analysis of agricultural field experiments (with discussion).” *Journal of the Royal Statistical Society: Series B*, 61, 691–746.
- Besag, J. and Kooperberg, C. (1995). “On conditional and intrinsic autoregressions.” *Biometrika*, 82, 733–746.
- Besag, J., York, J., and Mollié, A. (1991). “Bayesian image restoration, with two applications in spatial statistics.” *Annals of the Institute of Statistical Mathematics*, 43, 1–59.
- Borhani, N. O. (1965). “Changes and geographic distribution of mortality from cerebrovascular disease.” *American Journal of Public Health*, 55, 673–681.

- Botella-Rocamora, P., Martinez-Beneito, M. A., and Banerjee, S. (2015). “A unifying modeling framework for highly multivariate disease mapping.” *Statistics in Medicine*, 34, 1548–1559.
- Bradley, J. R., Wikle, C. K., and Holan, S. H. (2014). “Mixed effects modeling for areal data that exhibit multivariate-spatio-temporal dependencies.” ArXiv preprint, arXiv:1407.7479.
- Casper, M., Anda, R. F., Knowles, M., Pollard, R. A., and Wing, S. (1995). “The shifting stroke belt: Changes in the geographic pattern of stroke mortality in the United States.” *Stroke*, 26, 755–760.
- CDC (2003). “CDC/ATSDR Policy on Releasing and Sharing Data.” Manual; Guide CDC-02. Available at <http://www.cdc.gov/maso/Policy/ReleasingData.pdf>. Accessed June 30, 2015.
- Cressie, N. and Huang, H.-C. (1999). “Classes of nonseparable, spatio-temporal stationary covariance functions.” *Journal of the American Statistical Association*, 97, 1330–1340.
- El-Saed, A., Kuller, L. H., Newman, A. B., Lopez, O., Costantino, J., McTigue, K., Cushman, M., and Kronmal, R. (2006). “Geographic variations in stroke incidence and mortality among older populations in four US communities.” *Stroke*, 37, 1975–1979.
- Ergin, A., Muntner, P., Sherwin, R., and He, J. (2004). “Secular trends in cardiovascular disease mortality, incidence, and case fatality rates in adults in the United States.” *The American Journal of Medicine*, 117, 219–227.
- Gelfand, A. E. and Vounatsou, P. (2003). “Proper multivariate conditional autoregressive models for spatial data analysis.” *Biostatistics*, 4, 11–25.
- Gillum, R. F., Kwagyan, J., and Obiesesan, T. O. (2011). “Ethnic and geographic variation in stroke mortality trends.” *Stroke*, 42, 3294–3296.

- Gneiting, T. (2002). “Nonseparable, stationary covariance functions for space-time data.” *Journal of the American Statistical Association*, 97, 590–600.
- Howard, G., Howard, V. J., Katholi, C., Oli, M. K., Huston, S., and Asplund, K. (2001). “Decline in US stroke mortality: An analysis of temporal patterns by sex, race, and geographic region.” *Stroke*, 32, 2213–2218.
- Jin, X., Banerjee, S., and Carlin, B. P. (2007). “Order-free co-regionalized areal data models with application to multiple-disease mapping.” *Journal of the Royal Statistical Society, Series B*, 69, 817–838.
- Klebba, A. J. and Scott, J. H. (1980). “Estimates of selected comparability ratios based on dual coding of 1976 death certificates by the eighth and ninth revisions of the International Classification of Diseases.” *National Vital Statistics Reports*, 28.
- Knorr-Held, L. (2000). “Bayesian modelling of inseparable space-time variation in disease risk.” *Statistics in Medicine*, 19, 2555–2567.
- Knorr-Held, L. and Rue, H. (2002). “On block updating in Markov random field models for disease mapping.” *Scandinavian Journal of Statistics*, 29, 597–614.
- Kochanek, K. D., Murphy, S. L., and Xu, J. (2015). “Deaths: Final data for 2011.” *National Vital Statistics Reports*, 63, 3.
- Lisabeth, L. D., Diex Roux, A. V., Escobar, J. D., Smith, M. A., and Morgenstern, L. B. (2006). “Neighborhood environment and risk of ischemic stroke.” *American Journal of Epidemiology*, 165, 279–287.
- Martinez-Beneito, M. A. (2013). “A general modelling framework for multivariate disease mapping.” *Biometrika*, 100, 539–553.
- NCHS (2013). “Bridged-race population estimates: United States July 1st resident population

- by state, county, age, sex, bridged-race, and Hispanic origin.” http://www.cdc.gov/NCHS/nvss/bridged_race.htm. Accessed: June 30, 2015.
- Quick, H., Banerjee, S., and Carlin, B. P. (2013). “Modeling temporal gradients in regionally aggregated California asthma hospitalization data.” *Annals of Applied Statistics*, 7, 154–176.
- Quick, H., Carlin, B. P., and Banerjee, S. (2015a). “Heteroscedastic conditional autoregression models for areally referenced temporal processes for analysing California asthma hospitalization data.” *Journal of the Royal Statistical Society, Series C*, 64, 799–813.
- Quick, H., Holan, S. H., and Wikle, C. K. (2015b). “Zeros and ones: A case for suppressing zeros in sensitive count data with an application to stroke mortality.” *Stat*, 4, 227–234.
- Quick, H., Waller, L. A., and Casper, M. (2015c). “A nonseparable multivariate space-time model for analyzing county-level heart disease death rates for race and gender.” ArXiv preprint, arXiv:1507.02741.
- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields*. Boca Raton, FL: Chapman & Hall/CRC.
- Schieb, L. J., Mobley, L. R., George, M., and Casper, M. (2013). “Tracking stroke hospitalization clusters over time and associations with county-level socioeconomic and healthcare characteristics.” *Stroke*, 44, 146–152.
- Stein, M. L. (2005). “Space-time covariance functions.” *Journal of the American Statistical Association*, 100, 310–321.
- Tassone, E. C., Waller, L. A., and Casper, M. L. (2009). “Small-area racial disparity in stroke mortality.” *Epidemiology*, 20, 234–241.
- Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997). “Hierarchical spatio-temporal mapping of disease rates.” *Journal of the American Statistical Association*, 92, 607–617.

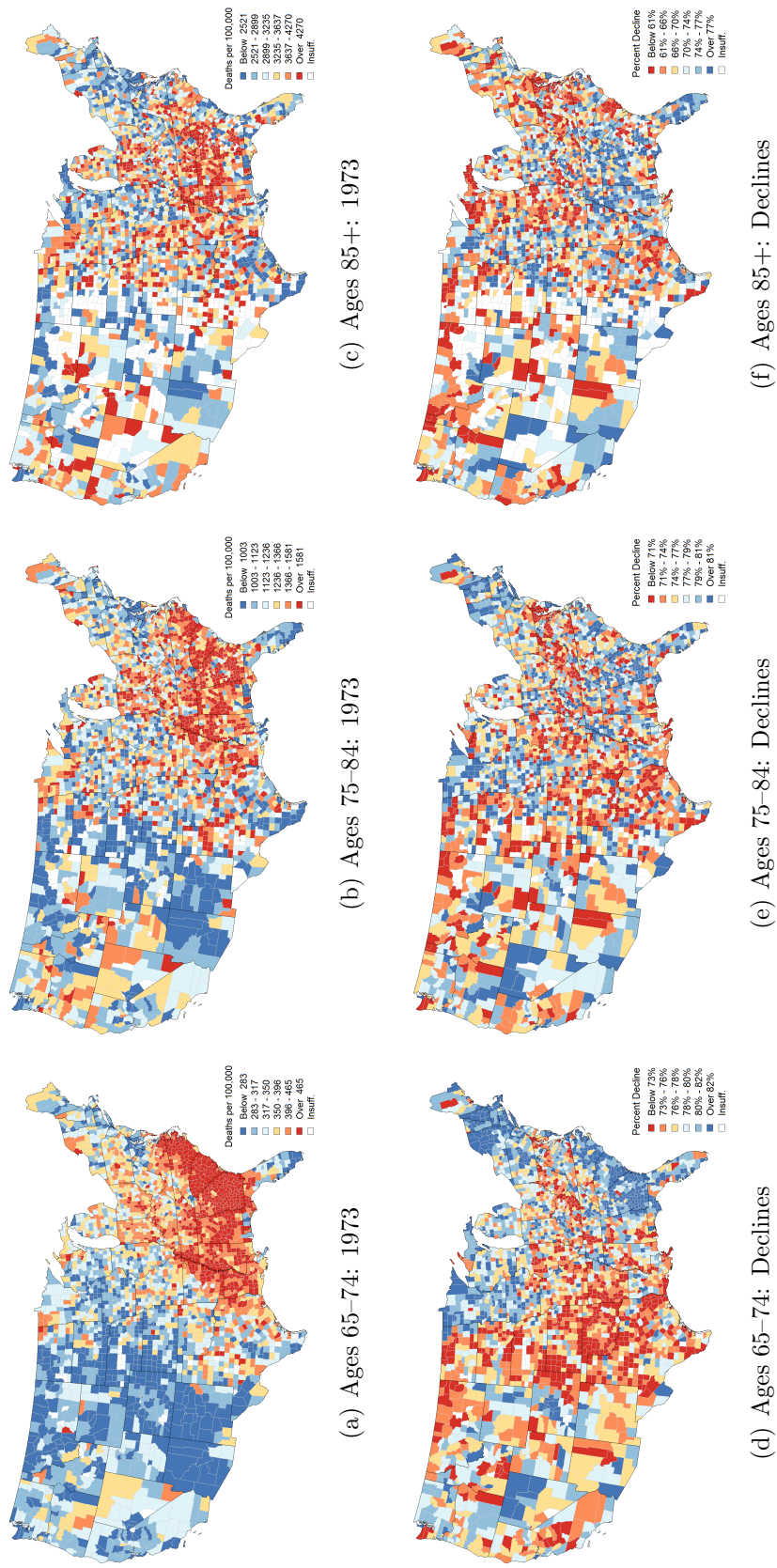


Figure 4: Maps of the stroke mortality rates at the beginning of the study period (1973) and the percent declines over the course of the study period. Note that estimates for counties with fewer than 100 people in an age bracket in 1973 are suppressed.